

AI for audience prediction and profiling to power innovative TV content recommendation services

Lyndon Nixon
MODUL Technology GmbH
Vienna, Austria
nixon@modultech.eu

Krzysztof Ciesielski
Genistat AG
Zürich, Switzerland
& ICS PAS, Warsaw, Poland
krzysztof.ciesielski@genistat.ch

Basil Philipp
Genistat AG
Zürich, Switzerland
basil.philipp@genistat.ch

ABSTRACT

In contemporary TV audience prediction, outliers are considered mere anomalies in the otherwise cyclical trend and seasonality components that can be used to make predictions. In the ReTV project, we want to provide more accurate audience predictions in order to enable innovative services for TV content recommendation. This paper presents a concept for identifying the source of outliers and factoring TV content categories and the occurrence of events as additional features for training TV audience prediction. We show how this can improve the accuracy of the audience prediction. Finally, we outline how this work could also be combined with AI-enabled audience profiling to power new content recommendation services.

CCS CONCEPTS

- Computing methodologies~Machine learning algorithms

KEYWORDS

Audience forecasting, prediction, predictive analytics, audience segmentation, viewer profiling, TV content recommendation

1 Introduction: The Need to Know Future Audiences

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

AI4TV '19, October 21, 2019, Nice, France

© 2019 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6917-6/19/10...\$15.00

<https://doi.org/10.1145/3347449.3357485>

TV channels benefit from being able to anticipate future viewer numbers. Private channels set advertising slot pricing according to the expected number of viewers of the programming into which the advertising is inserted. Public channels need to show they can fulfil the remit for which they are publicly funded, which typically includes maximizing the audience for programming which has a social or regional purpose. Public as much as private channels would value audience forecasts when making scheduling decisions or content purchasing/production decisions, by simulating the potential audience for different choices of which content is to be broadcast at which time.

ReTV (retv-project.eu) is an EU Horizon 2020 funded research project whose goal is to enable media organizations including broadcasters to optimize the publication of their media content across digital channels. Through analysis of the success of past content publication, we are building cross-channel prediction models to anticipate which (type of) content will potentially be most successful by channel and time in the future. This can inform organizational decisions regarding which content to publish as part of an optimized content publication strategy. This includes the creation of content summaries for different channels (e.g. social media video is generally shortened to the key segments to highlight to a user of that channel) as well as the recommendation of when and where to publish those summaries to optimize reach and engagement with the audience.

Generally, forecasting methods remove or ignore the significance of outliers in the time series data (see Section 2). ReTV has begun with improving the audience forecasting by combining the EPG data to add content categories as a new feature in the learning model (see Section 3). We then identified how outliers in audience figures are largely connected to event occurrences and hence began to collect relevant events to include them as a new feature in the learning model, so that we could take future events into account in the audience forecasting (see Section 4). We then tested collaborative filtering methods, traditionally used in recommendation, as a solution to feature-rich audience prediction and profiling (see Section 5). We test combining the various features – content, event and audience – and expect to find significant improvements in predicting future audiences for content as a result, which in turn can enable innovative TV content recommendation services (see Section 6).

2 Related Work

The prediction of future values of some continuous time-series data set is generally referred to as **forecasting** and has been applied in various domains. Time series are usually decomposed by prediction models like ARIMA into three components: trend, seasonality and remainder (also called ‘irregular’). An example decomposition is shown in Fig. 1.

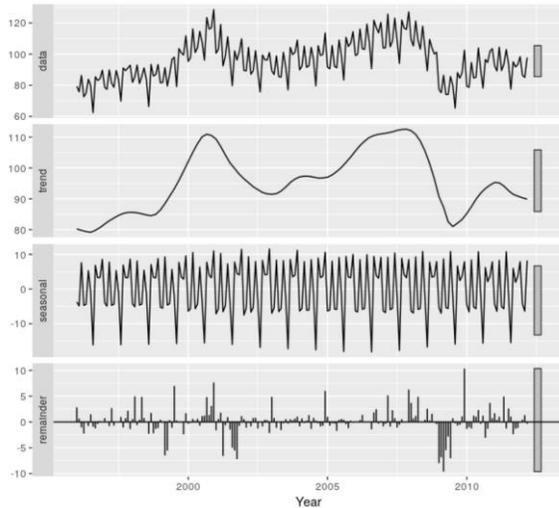


Figure 1: “The electrical equipment orders (top) and its three additive components” (Hyndman and Athanasopoulos, 2018).

The forecast is generated by projecting the identified trend and seasonal cycles of the data into the future while disregarding the irregular component, since it is by definition unpredictable. Outliers – the data points in the extremities of the remainder component - are a regular topic of discussion in statistical forecasting, since they are typically occurring in real world data and can indicate various states of relevance or irrelevance to the forecasting task. What is an outlier has to be specified explicitly or learnt from the data, i.e. determining the bounds of normality for the data measurement [1]. One rule of thumb is all data points three standard deviations away from the mean, referred to as the z -score = 3 [2]. This z -score (threshold) can naturally be modified to control how many data points are handled as outliers [3]. For example, outliers can be indicators of errors in the data measurement. As a result, forecasting models typically remove or reduce the effect of outliers as they would lead to less accurate prediction results. Yet there may be cases where “outliers are also regarded as noisy data, although they are actually extreme or exceptional, but correct, cases” [4]. Some work has considered that outliers may contain valuable information for prediction, e.g. abnormally low or high energy consumption in a building [5].

Regarding prediction of TV audiences, forecasting methods are applicable since TV programming can be both seasonal (e.g. summer vs winter schedules) and viewership follows identifiable trends (e.g. weekday ‘prime time’ in the early evenings) [6]. While linear forecasting models are predicting on the basis of weighted moving averages from the time series and “exclusively from the seasonality of past TV usage” [7], non-linear forecasting models can consist of several predictors. This is a newer area of

research since it makes use of AI techniques, e.g. features such as demographic/behavioural audience segments can also be added to the model and used in the prediction [8]. However we are not aware of any prior work using TV program topics or event occurrences as features for TV audience forecasting.

It has also been explored if there are correlations between other indicators and TV viewership. Typically there has been interest in the significance of social media activity, e.g. “for 18-34 year olds, an 8.5% increase in Twitter volume corresponds to a 1% increase in TV ratings for premiere episodes” [9]. The likes, shares and comments on TV show pages on Facebook or tweets and retweets on Twitter may be indicators of the show’s popularity and correlate to viewing figures of the next episode broadcast [10]. Such work to date appears to suggest that social media metrics are a viable feature for a prediction model, but do not address the issue of the bulk of TV programming which is not subject to a critical mass of social media discussion or content engagement.

Our work uniquely considers the content of the TV program and the occurrence of other events as features for predicting future audiences, learning also from outliers in the past data instead of smoothing them out.

3 Content-Based Audience Prediction

Our baseline audience prediction used random forest models on viewing numbers per TV channel. For training, we use data from Zattoo (an OTT TV provider in Switzerland and other European countries) that gives us the information about who watched which program on which channel and at what time (user IDs were anonymized prior to analysis and only aggregations of viewers were used in the forecasting). Real time data points (audience at every 5 minute time point in the last hour) were used to adjust predictions to most recent trends.

To analyze if the type of TV content being broadcast has an effect on audience, we used two sources of EPG metadata: (1) The first source contains an enhanced categorization of the programs (in particular, including different sport disciplines) and the start and end times are more accurate; (2) The second source contains a basic categorization (News, Documentary, TV Series, Entertainment, Kids, Movies, Sport) of the programs and the start and end times are approximate to about 5 minutes. Comparing results with both content feature sources allows us to verify whether (a) the model is flexible enough to use different kinds of attributes (b) how the information granularity affects the model quality.

To learn how the type of TV content affects the audience numbers, we took the past 5 months of audience data and matched it to the corresponding EPG data. We categorized the EPG data into five categories: sports (green), news (yellow), movies/TV series (blue), ads/promos (red) and other (black). Audiences numbers (dashed line) were smoothed to medians aggregated over channel, hourly and weekly seasonal variations. A sample plot of audience by TV content category is shown in Fig. 2.

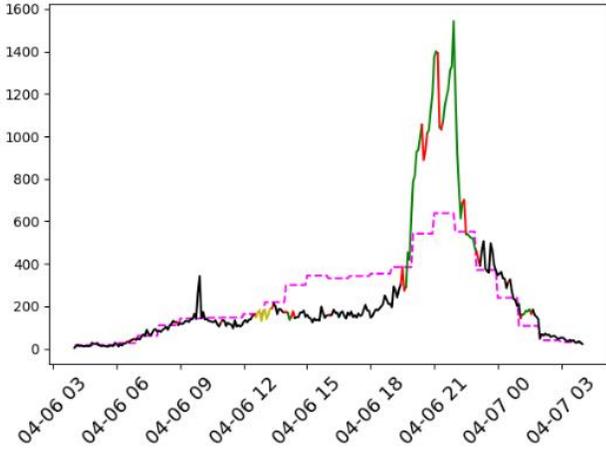


Figure 2: Plot of TV channel audience over 24 hours, colour coded by TV content category

Analysing the plots for all channels, we found that sport is related to most of the anomalies in audience figures. News is much less important. Longer ad breaks do lead to some audience erosion but it is also temporary. Channels that do not broadcast sport have very stable audience shapes for most of the time. Even the day-of-week (i.e. weekly) seasonality is not that important, just daily seasonality. The same holds for non-sport days on the other channels. This implies that the "typical" TV channel audience and its seasonality is enough to predict in many cases, without additional features. However, where a channel broadcasts a future content item which will cause an 'anomaly' in audience figures, as seen with live sports events, our classical prediction model could not predict this out-of-trend variation. So we decided to add the TV content categories as a feature (categorizing the EPG data for the next 24 hours of broadcast TV) to our prediction model to test if this could improve prediction.

4 Event-Based Audience Prediction

In various cases in TV audience data, external events (i.e. occurrences outside of the TV programming itself) can have an effect on viewing numbers alongside some TV-specific events (e.g. finale of a very popular program). For example, the Super Bowl is regularly among the most watched TV broadcasts in the USA. Events can be included in prediction models by using dummy variables with time-series multiple regression. The dummy variable will be binary, with value 1 = "yes" and 0 = "no" for whether the event occurred on that day or not (this is known as 'one-hot encoding' in machine learning). This avoids the simple removal of outliers, which may be associated with the presence of an event [10].

Firstly, since we do not want to build a prediction model with all possible events one-hot encoded – potentially introducing too many irrelevant features or accidentally determining correlations which do not hold – we ask which events actually are relevant to outliers in TV audience data. We took the audience data from Feb 16 to Oct 2, 2018 for several German and Swiss TV channels and chose several top channels from both countries: ARD, ZDF and PRO7 (in Germany) and SRF1 and SRF2 (in Switzerland). We used Anomaly Detection in SPSS. The initial threshold of three

standard deviations from the mean ($z\text{-score} = 3$) was too discriminatory and we settled on $z\text{-score} = 2$ for extracting anomalies in the data. This returned 25 data points in ZDF audience data instead of 4, for example (Fig. 3).

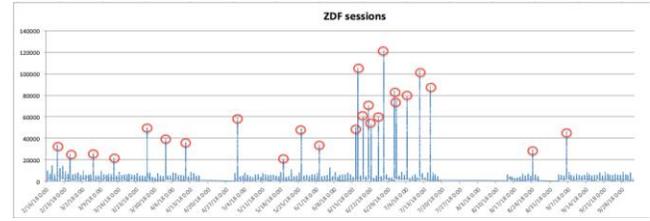


Figure 3: Outliers in ZDF audience data with $z\text{-score} = 2$

In Table 1, we summarize the results of looking at each anomaly for each channel and manually determining if they relate to (a) a TV specific event (like a series finale), (b) an external event broadcast on that channel (like live sports coverage), or (c) not explained. It can be seen that no anomaly was unexplainable. Only in the PRO7 case the anomalies occurred due to a TV-specific event, in fact they were the weekly broadcasts of "Germany's Next Top Model" which attracted a much higher audience than any other programming on that channel. The weekly repetition of these outliers could be used to learn that this is related more to the schedule of TV programming than to external events (which do not occur as regularly). For all other channels, we could explain all of the anomalies by events that occurred at that time and were broadcast on that channel, indicating both that outliers in audience data can be meaningful for prediction and that they need identification with events for prediction model learning.

Channel	Total # Anomalies	TV-only event	External event	Not explained
ZDF	25	0	25	0
ARD	18	0	18	0
PRO7	12	12	0	0
SRF1	1	0	1	0
SRF2	6	0	6	0

Table 1: Identification of relationship between Events and "Anomalies" in TV audience data

We also looked at the types of events associated with the anomalies. The vast majority were sports (most obviously, many FIFA World Cup games). In Germany only the Royal Wedding (Prince Harry and Meghan Merkle) and Eurovision Song Contest able to generate a similar spike in audience. In Switzerland, the SRF1 anomaly related to a Spring celebration parade in Zürich being broadcast, whereas all SRF2 anomalies were sports-related. Geographical location of the channel is also determinant of which events may cause anomalies, since all SRF anomalies except one related to events specifically involving Switzerland. We did not observe significant drops in audience on other channels at the same time, nor did we observe overall increases or decreases in

audience across all channels that could be related to an event (e.g. a public holiday). So our main focus in the event-based prediction will be on learning about past events' effect on TV audience and using this to predict TV audiences during future events. Having learnt which types of events specifically have been relevant to past TV audience figures' anomalies, we set up an event collection pipeline to build a Knowledge Base of future events of the same type. We used WikiData for an initial collection, identifying gaps in the event coverage such as individual sports matches. We added additional sports events using public calendars (iCal format) created by sports fan communities.

We extend our prediction model with event features, i.e. indicating the occurrence of an event during a certain time period. To capture that different events might affect TV audiences in different ways, we considered how to model a set of event features – each represented by an integer value – to represent significant differences between the considered events. Using a set of integers to represent a past event allows our model to learn how different events affect the audience and use this learning in prediction with future event representations. The features chosen for the model were:

1. Category of event (sport, entertainment, popular culture)
2. Sub-category (e.g. sport -> soccer, hockey, boxing etc.)
3. Location of the event (e.g. Germany, Switzerland)
4. Participants in the event (e.g. the two soccer teams or tennis players)
5. Stage (e.g. group match, quarter final, semi-final, final)

The third feature is restricted to the countries in which the measured channels broadcast, as it was observed that events involving the country attracted higher audiences than similar events not involving the country. This is used not only for events occurring within the country but also when the country is explicitly a participant (e.g. Switzerland national soccer team in a soccer match). The last two features are typically sports-specific and might receive null values for other events, but as the vast majority of events of interest are sports this is reasonable. It also worked well with the Eurovision Song Contest, capturing that there is a higher audience for the final compared to the semi-finals.

We trained our model on 4 weeks of past audience data aligned to events in our KB, focused on sports (there were very few events of other types in any case) and predicted for the next 24 hours based on the available EPG data. We observed much less improvement in prediction than with the content-based features. However the event-based prediction had more limitations. Firstly, we need to link future events in our KB to their broadcast on the TV channel, as the audience variation is dependent on the event being shown on the channel. This made it only possible to add event features for the next 24 hours of TV broadcast. We manually associated the events to TV programming in the EPG; an automatic approach would be dependent on the quality and completeness of the EPG program descriptions. Regarding the evaluation, it should also be noted that only 5% of TV content in

the EPG could be associated to a known event in our KB, meaning event features are less significant when evaluating over 24 hours of predicted audience.

5 Audience Profiling for Predictions

Another feature for audience forecasting is to segment the audience by viewing preferences [8]. Viewing preferences can be learnt directly from past audience data, i.e. preferences about what channels are watched on what day at what time. Assuming the preferences of the audience remains fundamentally the same, future audience can be predicted. In Section 3, we already noted that the type of content in the TV programming can also be a feature for a learning model, so that preferences represent what content is preferred by the audience. An advantage of the content-based audience profiling is that the preferences can be learnt across all TV channels rather than assuming every TV channel would have its own, individual and entirely separate viewing patterns. In other words, we can consider the prediction task to be to determine the likely percentage of the total audience (the sum of all individual viewers in our audience data) to watch a piece of TV content on a given channel at a given time.

We benefit from having access to data about individual viewers and their viewing sessions from Zattoo, anonymized and provided in a pre-aggregated form (we can not reconstruct a single viewers TV viewing). In our case, we want to forecast the total audience for a piece of TV content as the aggregation of audience segments learnt from the past audience data which would watch that content. The intention is not to segment viewers using sets of multiple, individual categorical interests (e.g. soccer fan, murder mystery fan etc.) which could be determined by an explicit collection of interests (e.g. as part of a viewer's profile). This could lead to wildly simplistic matching (e.g. TV program has category Sport, so is watched by all Sports fans). Rather, our model should determine implicitly the likelihood of each viewer to watch a particular category of TV content. This means every viewer implicitly belongs to some audience segment, i.e. a group of viewers who share a similar likelihood to watch TV across each category (we expect this to allow us to capture more complex relationships between viewing patterns such as there are Sports fans who are just as likely to watch News whereas others would only watch Sports). Now for a future broadcast time point, we can determine for a viewer if they are likely to be watching TV at that time and if so, given the content across channels, which content they are most likely to be watching. Finally, we use the implicit audience segments (fuzzy clusters learned by the model) to place different viewer groups on different channels, according to their interests.

We experimented with two modeling approaches:

1. Baseline model: standard collaborative filtering based on Non-Negative Matrix Factorization (NNMF) [12]. This model does not use any additional content- or event-related features. It just observes the interactions between users and content (TV programs).
2. Field-aware Factorization Machines (FFM) model [13]. This model is as an extension of the basic factorization model, and it allows us to test the additional features (Sections 3 and 4). Most

importantly, it models the interactions between the individual feature values as a dot-product of the associated weighted vectors.

Collaborative filtering is traditionally used in recommendation. Indeed, our starting point for using these approaches was to build a model of viewer preferences for content recommendation. For a given set of TV content options, we wanted to predict which TV content the viewer is most likely to watch. We have developed two TV content recommendation scenarios:

1. Content sWitch: we replace a „general audience“ program trailer in the TV stream with a trailer personalized to the user’s interests. The replacement is done in real-time in the IP stream and takes into account the lengths of the original and replaced trailer. This also necessitates content summarization to adapt the trailer duration, which is beyond the scope of this paper;

2. Chatbot 4U2: within a preferred messaging app (e.g. Telegram or Whatsapp) the user subscribes to a set of pre-selected content categories, interacting with a conversational chatbot. Links to video recommendations (e.g. snippets of last nights’ programming) are then delivered on a daily basis.

In the first scenario, we track audience behavior and train the model that learns the interaction patterns between individual users (and their associated attributes) and individual content pieces (and their associated features, such as category). In the second scenario, we only have a general, explicitly provided list of user categories, so the input information is less detailed and static (unless the user modifies his or her profile). This allows us to compare recommendation in these two contexts – one where the user can be identified by a log-in, the other where the user is not identifiable across sessions and we can only use the explicitly provided information.

The recommendation model is however also a prediction model, since it learns for any choice of TV content the likelihood of that content being watched by any audience segment. Here, rather than having multiple TV content items and a single audience segment (that the target viewer of a recommendation belongs to), we would consider a single TV content item and calculate the likelihood to watch across all audience segments.

It should be noted that the model is trained on a very sparse data (since every user’s viewing pattern covers only a small part of the total broadcast TV content) and it requires to fit a high number of parameters (each feature value, e.g. each user identifier, is associated with a vector of weights in a low-dimensional latent space). FFM models are also prone to overfitting and require (a) careful training with the evaluation and test datasets and optimization early-stopping if the train/evaluation metrics diverge, as well as (b) proper optimization of model hyperparameters. Due to the lack of space, we will not discuss it here in detail. We will just mention that we applied the Bayesian hyperparameter optimization approach¹. For the early-stopping, we used the options available in xLearn library² that provides fast

implementation of FFM models. We include a measurement of the strength of the interaction between the user and the content (i.e. our target value to be modeled). However, explicit feedback from the user regarding how satisfied/engaged he or she is with a given program is naturally missing from our viewing data. Therefore we based our model on the fraction of the program that the user watched. The assumption is that the more of the program the user has watched, the more relevant it was for them. On the other end, zapping between programs generates low target values that are considered as (implicit) negative feedback. We do not consider total watching duration since this would introduce bias and promote some content categories (e.g. movies are usually much longer than TV series or news).

We trained the prediction model with our categorized audience data and the implicit audience segments. There are two types of metrics that are involved in the recommendation model training:

1. metrics that are optimized during model fitting phase;
2. metrics that we use to evaluate when the model is good enough for our purposes.

For the model optimization, we used the standard metric provided by xLearn library, log-loss (equivalent to cross-entropy). It should be noted that our approach is based on providing a single content item recommendation to a given user. So the metric that optimizes only the top of the program ranking (instead of optimizing for all users data) is preferred here. In the future, we’d like to experiment with metrics such as WARP³.

For the model evaluation, presented below, we applied the three standard metrics:

- MAE (mean average error) between the observed targets and the model-predicted values,
- Pearson linear correlation between the observed targets and the model-predicted values
- Spearman rank correlations.

Especially the last metric is relevant for our scenarios, since we don’t care about absolute values predicted by our model, but rather about having relevant content pieces at the top of the recommendations ranking.

Table 2 shows that, compared to the baseline model (NNMF), a features-based model (FFM) could already establish significantly better results which were also moderately improved by the additional of coarse or detailed TV programming content categories.

¹ <https://github.com/fmfn/BayesianOptimization>

² <https://github.com/aksnzhy/xlearn>

³ https://lyst.github.io/lightfm/docs/examples/warp_loss.html

Model	MAE	Pearson corr.	Spearman corr.
NNMF	0.45	0.32	0.35
FFM (no EPG)	0.25	0.73	0.75
FFM (coarse EPG)	0.22	0.75	0.77
FFM (detailed EPG)	0.18	0.79	0.8

Table 2: Prediction model with TV content categorization as additional feature compared to the benchmark

We compared the baseline model (NNMF, not using any content or user features, beside their identifiers) with various variants of FFM models (using the attributes described in the previous section):

- NNMF model had MAE error 0.45 (the less the better) and rank correlation between recommended and actually watched programs only 0.35 (the more the better)
- The best FFM model (with hyperparameters optimization applied) without additional attributes achieved MAE 0.25 and rank correlation 0.75
- The model with just one additional attribute (EPG category for the TV content) achieved MAE 0.18 and rank correlation 0.8. It was slightly worse in case of the less-detailed EPG metadata: MAE 0.22 and rank correlation 0.77. It shows the importance of providing the model with a high-quality content metadata.

Interestingly, the model without additional attributes was also more prone to overfitting. It may be because the differences between training and evaluation datasets are driven by factors which are not explicitly observed in the data (i.e. content-related attributes). Models for the chatbot scenario (where the input is the set of interests explicitly provided by user, instead of user behavioral data – detailed interactions with content – as in case of Content sWitch) were – as expected – slightly worse than the model in the Content sWitch scenarios: MAE 0.23 (vs. 0.18 for Content sWitch) and rank correlation 0.72 (vs. 0.8 for the Content sWitch). Still, the results are much better than the baseline or the model without any additional attributes provided. In both use case scenarios, the model took advantage of the interactions between the content features and the (implicit or explicit) user interests.

Future work is to test the recommendation model for audience prediction, aggregating audience segments that are most likely to watch a piece of future TV content. Planned model improvements include:

- using WARP instead of log-loss optimization – this will focus on the top of the recommendation ranking, instead of the complete ranking;

- testing if explicit audience segmentation improves the model (e.g. k-means clustering of viewers by watching preference) compared to the current, implicit fuzzy approach;
- including temporal features in the model (with the assumption that the current user session provides a better context for recommendations than previous sessions).

6 Future Work and Conclusions

The currently tested prediction model uses the following attributes:

- user: identifier, behavioral profile (the percentage of the time spent on each individual EPG category, which can be also viewed as implicit fuzzy segmentation of users)
- program: identifier, main EPG category (e.g. sport), detailed EPG genre (e.g. discipline) (actual values depending on the EPG metadata provider)
- events: stage, participants country, participant name

The prediction modeling could still be improved as results vary greatly between EPG categories. In general, the model works best for the popular categories such as sport, since we also have most training data for such categories. In parallel, we work on extending the additional attributes of events. As noted in Section 4, audience data contains anomalies which can be to a large extent attributed to events (sport events in particular) broadcast on TV. The big advantage of the FFM model is that it is able to model the interactions between the various feature values, so it automatically learns, e.g. that sport events mostly affect the behaviour of a sport-predisposed audience segment. Similar to users and content, we add an event identifier and set of event features (the more detailed the better, including temporal and geographical features of relevance). Later we are interested in also adding:

- behavioural viewership patterns (hours of the day, days of the week) in order to be able to find not only a proper content but also optimal engagement time, and
- more advanced content features such as face detection with Deep Neural Networks. It could help to fine-grain user preferences even more, capturing user interest in a given TV presenter or an actor.

In conclusion, we have learnt that in audience prediction we can improve forecasts by taking into the account the category of the TV content. While we have seen in the data how specific events cause significant anomalies in audience trends, we are still learning how best to incorporate event knowledge into our prediction model. The sparsity and irregularity of events as part of overall audience measurement is a limitation. We also can implicitly segment the (actual or predicted) audience and use this in TV content recommendation. We found that the TV program

category and overall content popularity as learnt by the recommendation model is even more important than an individual user profile. This may be considered a positive aspect of the model, since for a new user it allows to partially alleviate the cold-start problem (i.e. to recommend generally popular content rather than a random one, and iteratively learn the user preferences). We are now testing the accuracy of this recommendation model in predicting future audiences by aggregating the audience segments likely to watch a piece of future TV content. In general, we have found that AI models with additional features do work better but in terms of feature selection, content-based features have proven more effective to date compared to audience-based and then to event-based. The predictive analytics will be used in the ReTV project to provide tools for media organisations to help them publish the right content on the right channel at the right time. Two scenarios demonstrate how AI enabled audience prediction and profiling can power new innovative TV content recommendation services for TV viewers.

ACKNOWLEDGMENTS

This work is funded by the EU Horizon 2020 research and innovation programme under grant agreement H2020-780656 ReTV. MODUL Technology thanks Bachelors student Sarah Fuchs whose thesis research on anomalies in TV audience data contributed significantly to Section 4. GENISTAT thanks Zattoo AG for access to program metadata and anonymized audience data.

REFERENCES

- [1] Su, X. & Tsai, C.L. (2011). Outlier Detection. WIREs Data Mining and Knowledge Discovery, vol. 1, pp. 261–268.
- [2] Barnett, V. & Lewis, T. (1994). Outliers in Statistical Data, 3rd edition. John Wiley & Sons.
- [3] Chandola, V., Banerjee, A. & Kumar, V. (2009). Anomaly Detection: A Survey. ACM Computing Surveys, 41 (3), Article 15.
- [4] García, L. P. F., de Carvalho, A. C. P. L. F & Lorena, A.C (2013). Noisy Data Set Identification. Hybrid Artificial Intelligent Systems, pp. 629–638
- [5] Li, X., Bowers, C. P., & Schnier, T. (2009). Classification of energy consumption in buildings with outlier detection. IEEE Transactions on Industrial Electronics, 57(11), 3639-3644.
- [6] Danaher, P. J., Dagger, T .S., Smith, M. S. (2011). Forecasting Television Ratings. International Journal of Forecasting, vol. 27, pp. 1215–1240.
- [7] Weber, R. (2002). Methods to Forecast Television Viewing Patterns for Target Audiences. Communication Research in Europe and Abroad – Challenges of the First Decade. Berlin: DeGruyter.
- [8] Meyer, D., & Hyndman, R. J. (2006). The accuracy of television network rating forecasts: The effects of data aggregation and alternative models. Model Assisted Statistics and Applications, vol. 1(3), pp. 147–155.
- [9] Nielsen (2013). New Study Confirms Correlation Between Twitter and TV Ratings. Retrieved from: <https://www.nielsen.com/us/en/insights/news/2013/new-study-confirms-correlation-between-twitter-and-tv-ratings.html>
- [10] Hsieh, W.-T., Chou, S.-C., Cheng, Y.-H. & Wu, C.-M. (2013). Predicting TV Audience Rating with Social Media. IJCNLP 2013 workshop on Natural Language Processing for Social Media (SocialNLP), p. 1-5.
- [11] Hyndman, R. J. & Athanasopoulos, G. (2018). Forecasting: Principles and Practice, 2nd Edition. OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 12 March 2019.
- [12] Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. Journal of machine learning research, 5(Nov), 1457-1469.
- [13] Juan, Y., Zhuang, Y., Chin, W. S., & Lin, C. J. (2016, September). Field-aware factorization machines for CTR prediction. In Proceedings of the 10th ACM Conference on Recommender Systems (pp. 43-50). ACM.